REVIEW
Glycosylation of *Pichia pastoris*-derived proteins

Roger K. Bretthauer[1] and Francis J. Castellino

Department of Chemistry and Biochemistry and the W. M. Keck Center for Transgene Research, University of Notre Dame, Notre Dame, IN 46556, U.S.A.

The *Pichia pastoris* system for expression of heterologous recombinant proteins is being used increasingly because of the large yields of properly folded proteins that result and the ease of scaling preparations into large-biomass fermentors. Another advantage of this system centres on the type of glycosylation that results, generally yielding protein-bound oligosaccharides that are of much shorter chain length than found in *Saccharomyces cerevisiae*. This review is a summary of the current state of knowledge of glycosylation of proteins in this methylotrophic yeast.

Introduction

The *Pichia pastoris* expression system has been used extensively over the past decade for generation of recombinant proteins, and the nature of the processing that occurs in yeast has become a topic of interest. Yeast can carry out glycosylation of the amide nitrogen of asparagine residues in a protein when found in the consensus sequence Asn-Xaa-Thr/Ser, providing N-linked glycosylation. Also, glycosylation of hydroxy groups of threonine and/or serine residues in proteins occurs in yeast cells, yielding O-linked type of glycosylation [1]. As delineated mainly in *Saccharomyces cerevisiae*, protein glycosylation involves the following summarized events. N-linked oligosaccharides originate from an oligosaccharide ($Glc_3Man_9GlcNAc_2$) that is assembled on dolichol (pyrophosphate) in the endoplasmic reticulum (ER), and that is transferred to the appropriate Asn of the nascent protein in a co-translational event [2,3]. This is a common eukaryotic-cell-type glycosylation pathway for N-linked glycoproteins. However, differences in subsequent processing of the newly glycosylated protein occur in yeast as compared with higher eukaryotic cells (plants, insects and higher animals). In most yeasts that have been studied, the three glucose residues and one specific α-1,2-linked mannose residue are removed by specific glycosidases in the ER [4], giving a protein N-linked $Man_8GlcNAc_2$ core structure that is further processed in the Golgi complex [5,6]. This involves addition of a mannose residue that is linked α-1,6 to the α-1,3-linked mannose residue in the Manα-1,3Manβ-1,4GlcNAc inner-core sequence. The 1,6-linked residue can then be elongated to an α-1,6-linked backbone (50–100 residues) that can be branched with α-1,2-linked di- or tri-saccharides of mannose that in turn can be capped with α-1,3-linked mannose units. An additional processing event that occurs in most yeasts is the addition of phosphate, in the form of mannose 1-phosphate, to specific α-1,2- or α-1,6-linked mannose residues of the core or the side chain of N-linked oligosaccharides, forming the acidic phosphodiester component found in many yeast glycoproteins [7]. Complex-type oligosaccharides containing sialic acid, galactose, fucose and *N*-acetylgalactosamine are not found on *S. cerevisiae*-derived, and most other yeast-derived, glycoproteins. Therefore, the assembly and processing of only the high-mannose type of N-linked oligosaccharides on the expressed recombinant protein must be taken into account when considering the structure and the subsequent use (function) of the glycoprotein. Whereas cell-wall mannans from some strains of *P. pastoris* have been found to contain β-1,2-linked mannose residues, such as present in the structure Man(β-1,2)Man(β-1,2)Man(α-1,2)Man, β-linked mannose residues have not been reported in recombinant glycoproteins from *P. pastoris* [8,9]. In addition to the N-linked oligosaccharides, the presence of O-linked oligosaccharides of mannose is common in most yeasts [10]. The biosynthetic process is unique compared with animals, in that the mannose residue that is α-glycosidically linked to the hydroxy group of a serine or threonine residue of the protein is derived in the ER from a dolichol phosphate carrier, dolichyl phosphate mannose, rather than from a sugar nucleotide in the Golgi complex. These short oligosaccharides can also be phosphorylated and thus contribute to the acidity and perhaps other functions of the glycoprotein [7].

The status of glycosylation of recombinant proteins from *P. pastoris* has been presented in earlier reviews of 1992

Abbreviation used: ER, endoplasmic reticulum; endo H, endo-β-*N*-acetylglucosaminidase H; PNGase F, peptide N-glycohydrolase; DE-TOF-MALDI MS, delayed-extraction time-of-flight matrix-assisted laser desorption ionization MS.

[1] To whom correspondence should be addressed.

[11] and 1993 [12], and briefly in a recent review [1]. The purpose of this current review is to summarize the information that has been acquired over the past 10–12 years on the structures of oligosaccharides of recombinant proteins from *P. pastoris*. The large body of information available from previous studies with *S. cerevisiae* has greatly aided in understanding and elucidating these structures.

Gel-shift assays

Making use of the knowledge that only high-mannose-type oligosaccharides are found on yeast glycoproteins has afforded a relatively simple experimental method for determining whether N-glycosylation has occurred on the recombinant protein. The availability of enzymes that recognize the high-mannose core structure and catalyse hydrolysis of the GlcNAcβ–1,4GlcNAc linkage in the core, resulting in release of the oligosaccharide, generally results in sufficient decrease in apparent mass that can be detected using SDS/PAGE. Endo-β-*N*-acetylglucosaminidase H (endo H) is one such enzyme that has been used extensively for this purpose. The other type of enzyme utilized for release of the N-linked oligosaccharide is not specific for high-mannose or complex-type structures, and catalyses hydrolysis of the GlcNAcβ-Asn linkage. This enzyme is peptide-N-glycohydrolase F (glycopeptidase F or PNGase F), and it has been utilized extensively for demonstrating shifts in SDS/PAGE mobilities as an indication of N-glycosylation. Reports where N-glycosylation has been demonstrated only by use of either endo H or PNGase F in a gel-shift assay are listed in Table 1. The observed differences in apparent masses before and after endo-enzyme treatment are shown.

N-linked oligosaccharides from *P. pastoris*

Much of the current knowledge of the structures of the asparagine-linked oligosaccharides that are assembled on recombinant proteins from *P. pastoris* has been obtained with recombinant *S. cerevisiae* invertase. Expression in *P. pastoris* of the *SUC2* gene for this protein gave a secreted glycosylated invertase that was shown to contain less carbohydrate, due to shorter asparagine-linked oligosaccharides, than the invertase secreted by *S. cerevisiae* [13]. The majority (85%) of the endo H-released oligosaccharides from the *P. pastoris*-expressed invertase were of $Man_{8-14}GlcNAc$ sizes, with the remainder being much larger ($Man_{>30}GlcNAc$ size) but still shorter than the long oligosaccharide chains ($Man_{>50}GlcNAc$) found on the native *S. cerevisiae* invertase [14]. A detailed ^{1}H-NMR study of the structure of the different-chain-length oligosaccharides that were released by endo H from the heterologous invertase

Table 1 Endoglycosidase/gel shifts as evidence for glycosylation

Endo F, endo-β-*N*-acetylglucosaminidase F.

Protein	Change in molecular mass (kDa)	Glycosidase	Reference
S. cerevisiae invertase	100–140 to 58	Endo H	[13]
HIV envelope protein	200 to 59	Endo H	[36]
Boophilus microplus Bm86 antigen	90–100 to 70	Endo H	[37]
Bacillus amyloliquefaciens α-amylase	Smear to 60	Endo H	[38]
Bovine enterokinase catalytic subunit	46 to 33	Endo H	[39]
Rabbit angiotensin-converting enzyme	90 to 70	PNGase F	[40]
S. cerevisiae α-1,2-mannosyltransferase (Ktr1p) soluble domain	Reduced by 2.5	Endo H	[41]
Bovine herpes virus type-1 glycoprotein D	68 to 39	Endo H	[42]
Human prolyl 4-hydroxylase α-subunit	Three bands to one	PNGase F	[43]
Mouse gelatinase B	110 to 95	PNGase F	[44]
Human angiostatin (Pg K1–4)	51 to 49	Endo H	[45]
Bovine opsin	Reduced	Endo H or PNGase F	[46]
Human costimulatory B7-1 and B7-2, and B7 receptor CTLA-4	Reduced	PNGase F	[47]
Golgi sialoglycoprotein MG160	140 to 130	Endo H or PNGase F	[48]
Human α_1-antitrypsin	52 to 46	Endo H	[49]
Plasminogen activator inhibitor-1	84–100 to 50	PNGase F	[50]
Human placental alkaline phosphatase	62 and 65 to 58	Endo H	[51]
Mouse 5-hydroxytryptamine receptor	61 to 55 and 38	Endo H or PNGase F	[52]
Bla g 4 cockroach allergen	23 to 18 and 19	Endo F/PNGase F	[53]
Human tissue factor	37–45 to 33	Endo H	[54]
Human CD5-110	22.9 to 14.8	Endo H	[55,56]
Trypanosoma cruzi acid α-mannosidase	65 to 57 and 46	Endo H or PNGase F	[57]
Porcine follicle-stimulating hormone	23 to 15.5 and 24 to 18	Endo F/PNGase F or endoH	[58]
Carcinoembryonic antigen domains N and A3 fusion	37 to 24	PNGase F	[59]
Type-II activin receptor extracellular domain	30 to 25	Endo H	[60]
Bovine herpes virus type-1 glycoprotein bovine interleukin-6 chimaera	100 to 59.5	Endo H	[61]

and fractionated on BioGel P-4 was conducted [15]. The major conclusions were: the $Man_8GlcNAc$ has the normal yeast processing structure that lacks the α-1,2-Man on the middle arm of the $Man_9GlcNAc$ in the dolichol-linked precursor oligosaccharide; 95% of the $Man_9GlcNAc$ has an α-1,6-linked Man on the α-1,3-Man arm of the $Man_8GlcNAc$; 80% of $Man_{10}GlcNAc$ has an α-1,2-Man on the new α-1,6-Man of $Man_9GlcNAc$, 12% has an α-1,2-Man on other α-1,2-Man termini, and 8% has an α-1,6-Man on the new α-1,6-Man of $Man_9GlcNAc$; the $Man_{11}GlcNAc$ has four structural isomers, the major one (82%) being the major $Man_{10}GlcNAc$ with an additional α-1,2-Man on one of the three α-1,2-Man termini, and the other three structures having an α-1,2-Man on either one of the two new α-1,6-Man residues or on the α-1,2-Man terminus of the α-1,6-Man arm of the $Man_{10}GlcNAc$. Significantly, no new α-1,3-Man linkages were observed in these structures, implying that *P. pastoris* does not have, or does not utilize, α-1,3-mannosyl transferases for the processing of glycoproteins. Assuming that the dolichol-linked oligosaccharide in *P. pastoris* has the same $Man_9GlcNAc$ structure as found in other eukaryotic cells, only the specific α-1,2-mannosidase (for generating the yeast protein-linked $Man_8GlcNAc_2$), α-1,2-mannosyl transferase(s) and α-1,6-mannosyl transferase(s) would be required for processing of newly synthesized glycoproteins. The lack of α-1,3-Man linkages in the recombinant invertase oligosaccharides was later confirmed by studies showing that membranes from *P. pastoris* cells did not exhibit α-1,3-mannosyl transferase activity with GDP-Man as mannosyl donor and appropriate mannose-containing oligosaccharides as acceptors, whereas α-1,6- and α-1,2-mannosyl transferase activities were present [16].

A comparative study of N-linked glycosylation on six different recombinant proteins secreted from *P. pastoris* revealed some differences in chain lengths of the PNGase F-released oligosaccharides [17]. The relatively short $Man_8GlcNAc$ and $Man_9GlcNAc$ oligosaccharides were most commonly observed. Shorter oligosaccharides ($Man_7GlcNAc$) were abundant only on yeast (*S. cerevisiae*) invertase, and longer oligosaccharides ($Man_{10}GlcNAc$ up to $Man_{14}GlcNAc$) were abundant only on fungal (*Mucor pusillus*) aspartic protease and bovine enterokinase catalytic subunit. The other three proteins, fungal (*Penicillium minioluteum*) dextranase, bacterial (*Bacillus licheniformis*) α-amylase and tick (*Boophilus microplus*) gut antigen, contained predominantly $Man_8GlcNAc$ and $Man_9GlcNAc$ oligosaccharides. In this latter study, little differences in the size of the oligosaccharides could be correlated to whether the recombinant protein was prokaryotic or eukaryotic, composition of the growth medium or culturing in shake flask or fermentor. As the same integrative expression vector containing the yeast invertase signal sequence and the glyceraldehyde-3-phosphate dehydrogenase terminator sequence was used for expression of all proteins but the enterokinase, differences in transcriptional regulation would be considered to be minimal. However, the data suggest an inverse correlation of the oligosaccharide size with the number of potential N-linked sites and the expression level. It is suggested that the flow of newly synthesized glycoprotein through the secretory pathway might be so great as to not allow sufficient or proper processing of the oligosaccharide chains. In this case, use of weaker promoters other than the strong AOX1 promoter might allow enhanced glycosylation processing.

The N-glycosylation by *P. pastoris* of Asn-Xaa-Thr/Ser sites in prokaryotic proteins that are normally not glycosylated in their native hosts has been studied and was clearly demonstrated with the α-amylase of *Bacillus licheniformis* [18]. The protein contains six N-linked sites that could potentially be glycosylated. A decrease in mass from approximately 90 to 60 kDa after endo H treatment of the expressed protein indicated that substantial N-glycosylation had occurred. Location of the glycosylated sites in the recombinant protein was based on trypsin digestion before and after treatment with endo H, and subsequent fractionation of the peptides by HPLC. Monitoring of the peptides with concanavalin A and wheatgerm agglutinin lectins (as their horseradish peroxidase conjugates), along with peptide sequencing and MS, allowed assignment of peptides that did contain carbohydrate as the original high-mannose oligosaccharide or as the residual GlcNAc after endo H treatment. Although the structures of the oligosaccharides were not investigated, five of the six potential N-sites were assigned as being glycosylated by these techniques.

Glycosylation of other recombinant fungal enzymes from *P. pastoris* has been reported. An *Aspergillus fumigatus* catalase gene has been expressed in *P. pastoris* [19]. The protein has four potential N-linked sites, and glycosylation of the recombinant protein was indicated by reaction with concanavalin A–peroxidase conjugate on a gel blot. No other structural information on the oligosaccharides was presented. Expression of glucoamylase of *Aspergillus awamorii* yielded a recombinant protein that was analysed for neutral sugar with the phenol/sulphuric acid colorimetric procedure, and for mass by SDS/PAGE and MS [20]. The results were consistent with the recombinant enzyme containing 96 neutral hexose molecules. The structures of the carbohydrate units were not further investigated.

Modification of the Asn-linked glycosylation pathway in yeast

Attempts to genetically modify glycosylation processing in *P. pastoris* were carried out [21]. Recombinant strains of *P.*

pastoris containing an integrated cDNA for either secreted haemagglutinin or influenza neuraminidase were further transformed with a full-length cDNA for α-1,2-mannosidase from *Trichoderma reesei* or with a fusion cDNA of the catalytic domain of the *T. reesei* α-1,2-mannosidase and the ER retention signal for the *S. cerevisiae* MNS1 protein (specific ER-processing α-1,2-mannosidase). Oligosaccharides from neuraminidase secreted by cells containing the heterologous α-1,2-mannosidase were shown to be partially reduced in size from $Man_{<14}GlcNAc$ observed in control cells to $Man_{5-8}GlcNAc$, implying that some additional trimming of mannose residues was occurring with the α-1,2-mannosidase. As the *T. reesei* mannosidase was observed to be secreted from the *P. pastoris* cells, its catalytic domain was anchored to the ER by using the fusion cDNA containing the *S. cerevisiae* retention signal. Neuraminidase secreted by these *P. pastoris* cells did not contain the shorter $Man_{<14}GlcNAc$ oligosaccharides, and secreted haemagglutinin oligosaccharides contained additional short $Man_{5-9}GlcNAc$ and longer hyperglycosylated components. It was suggested that the presence of the heterologous α-1,2-mannosidase allowed formation of better substrates for mannosyltransferases, and thus the formation of larger mannosylated oligosaccharides. Further investigations of this type could result in expression of recombinant proteins that would not be compromised in structure or function by the large mannosylated side chains.

Identification of sites of N-linked glycosylation in glycoproteins

The location in the polypeptide chain of the Asn-Xaa-Thr/Ser glycosylation site, and the presence of tandem or overlapping sites, were shown to have significant effects on the size of the oligosaccharide chains in recombinant single-chain Fv antigen-binding protein (scFv) [22]. Having a single site near (1–5 residues) the C-terminal end of the protein resulted in only core ($Man_{8-14}GlcNAc$) glycosylation, with high efficiency (95%) for the five-residue-spaced site. The presence of tandem sites near the C-terminal end induced hyperglycosylation (increase in mass by 9–15 kDa), with three tandem sites being completely hyperglycosylated. Interestingly, the presence of a single two-site overlapping sequence (NNTT) adjacent to the C-terminal serine residue resulted in core glycosylation (minor) and hyperglycosylation (major). It was therefore suggested that the sequence NNTT delineates a minimal sequence for one form of hyperglycosylation. But it is also cautioned that in these experiments the N-glycosylation sites are engineered to be very near the C-terminus of the recombinant protein, a glycosylation site that is found rarely in native proteins.

Characterization of glycosylated recombinant proteins from *P. pastoris*

Use of expression vectors for *P. pastoris* that contain the yeast α-mating-factor signal sequence (such as pPIC9) can give results on glycosylation that must be interpreted carefully because of the potential N-glycosylation sites in this secretion signal sequence. An example of this is the report on expression of human procarboxypeptidase A2 (proCPA2) in *P. pastoris* [23]. The pPIC9 vector was used to construct the expression vector for secretion of the recombinant protein. On SDS/PAGE, a major 45-kDa secretion product corresponding to proCPA2 was obtained, and also a smear of higher-molecular-mass immunopositive material that was reduced to 50–60 kDa mobility with PNGase F treatment. This latter protein corresponded to αMF-proCPA2, and was interpreted to result from incomplete *in vivo* processing of the glycosylated construct by the *P. pastoris* KEX2 protease. N-terminal amino acid sequencing of recombinant proteins is generally required and utilized to ascertain that proper protease processing has occurred.

Proteins that are expressed in glycosylated forms by *P. pastoris* are suitable for crystallization and subsequent structural analysis. The glycosylated $\alpha_E C$ protein domain of human fibrinogen-420 was expressed in *P. pastoris* and shown to contain oligosaccharides released by PNGase F of sizes $Man_{6-13}GlcNAc_2$ and minor amounts of phosphorylated components ($Man_2PGlcNAc_2$) [24]. A segment of the expressed domain containing residues 647–847, including the N-glycosylation site at Asn^{667}, was crystallized for structural analysis. Electron density at Asn^{667} in the resulting electron-density map was attributable to the N-linked GlcNAc of the carbohydrate cluster [25].

A more detailed account of the ability to crystallize *P. pastoris*-expressed and -glycosylated proteins has been reported for human pancreatic α-amylase [26]. It was shown by electrospray ionization MS that the protein contained one N-linked oligosaccharide of sizes $Man_{8-11}GlcNAc_2$, and in this glycosylated form was not crystallizable. However, after treatment with endo-β-*N*-acetylglucosaminidase F, the protein was crystallized and the electron density at Asn^{461} shown to fit a GlcNAc residue. Important and interesting results of this glycosylation-site assignment were that Asn^{461} is not part of an Asx or β-turn that favour glycosylation, and that the cysteine residue in the Asn^{461}-Cys-Thr consensus sequence is in disulphide linkage with Cys^{450} in the native protein. Few other proteins have been found to have this structure around a glycosylation site.

Structures of the various glycoforms of the recombinant kringle 2 domain of tissue-type activator ($r\text{-}K2_{tPA}$) [27] have been analysed in our laboratories [28,29]. Release of

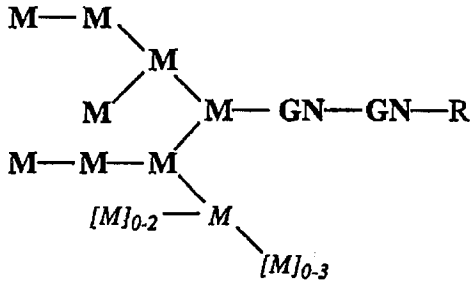

Figure 1 An oligosaccharide structure accounting for the hydrodynamic sizes of $Man_{9-12}GlcNAc_2$ and their α-1,2-mannosidase-catalysed digestion products

Residues shown in regular type are derived from the $Glc_3Man_9GlcNAc_2$ oligosaccharide that is assembled through the dolichol pathway, transferred to the protein, and further processed by glucosidases I and II, and an ER-localized α-1,2-mannosidase. The resulting $Man_8GlcNAc_2$ core structure has been retained in all of the proposed structures. The other mannose residues present in the oligosaccharides (italicized) are derived from later processing reactions. Glycosidic linkages from the Man-GlcNAc-GlcNAc-R component of the core structure are displayed as —(α-1,2)/(α-1,3) and\(α-1,6) linkages. M, mannose; GN, GlcNAc. This Figure is reproduced from [28] © 1997 Portland Press on behalf of IUBMB.

the oligosaccharides from the protein by PNGase F or by hydrazinolysis gave identical profiles on subsequent chromatography of the fluorescently labelled products, indicating that only N-linked saccharides were present. Fractionation by gel filtration on BioGel P-4 or by HPLC on amino-silica revealed a series of oligosaccharides that corresponded to $Man_{8-18}GlcNAc_2$, with the major components (75%) being $Man_{9-12}GlcNAc_2$. Structures of the major individual oligosaccharides were analysed by using a combination of α-mannosidases (non-specific jack bean, 1,2-specific *Aspergillus saitoi* and 1,6-specific *Xanthomonas manihotis*), gel filtration and delayed-extraction time-of-flight matrix-assisted laser desorption ionization MS (DE-TOF-MALDI MS). Results of these analyses could be accounted for by a $Man_8GlcNAc_2$ core structure (yeast type) having an extension on the core α-1,3-Man arm of up to four α-1,6-Man units and up to two α-1,2-Man units on either the $Man_8GlcNAc2$ core or the α-1,6-linked side-arm extension. No α-1,3-Man glycosidic linkages were observed (or required to account for the experimental data). The types of structures that can be observed are summarized in Figure 1. All of these results are in agreement with earlier structural and enzymic studies pertaining to *P. pastoris*, and with a large body of knowledge pertaining to protein glycosylation in *S. cerevisiae*.

O-glycosylation of proteins from *P. pastoris*

Evidence for O-glycosylation of recombinant proteins expressed in *P. pastoris* is limited. An early study involving mild alkaline treatment of glycoproteins of glycoproteins from the periplasmic space and from secretions of cells expressing heterologous invertase indicated by gel filtration that O-linked oligosaccharides were present but were not major components as compared with the N-linked oligosaccharides [14]. For recombinant human lysosomal α-mannosidase, some carbohydrate was resistant to removal by endo H and N-glycanase, as indicated by retention of staining with *Galanthus nivalis* lectin; this was attributed to the presence of O-linked mannose structures [30]. A human single-chain urokinase-type plasminogen activator protein without an N-glycosylation site (Asn → Gln mutation) was expressed and secreted, and was shown to stain on a gel blot for carbohydrate [31]. It was therefore concluded that the carbohydrate was O-linked to this recombinant protein.

Two studies having direct structural evidence for O-mannosylation of expressed proteins in *P. pastoris* have since appeared. One was conducted with recombinant forms of mouse gelatinase B [32]. Several glycosylated truncated variants of this protein were expressed, ranging in size from 112 to 60 kDa on gel electrophoresis. Variants (75 and 60 kDa) lacking N-glycosylation sites were still stained with *G. nivalis* lectin, suggesting that O-linked saccharides were present. Release of bound saccharides by hydrazinolysis yielded oligosaccharides ranging in size by HPLC from mono- to pentasaccharides, and which were completely degraded to mannose by jack-bean α-mannosidase. These results are consistent with O-linked saccharides of mannose being assembled on serine or threonine residues of the newly synthesized recombinant protein. The second study of O-glycosylation in *P. pastoris* was carried out in our laboratory with recombinant kringle 1–4 domain of human plasminogen [33]. Release of saccharides from the expressed protein was accomplished by hydrazinolyis and with alkaline β-elimination. Both procedures yielded oligosaccharides up to pentamer in size (gel filtration) that were completely degraded to monomer by jack-bean α-mannosidase. Treatment with a specific α-1,2-mannosidase degraded essentially all of the saccharides to monomer, indicating that no α-1,3-linked mannose units were present in the O-linked oligosaccharides.

Phosphorylation of oligosaccharides in *P. pastoris*

The presence of phosphate in *S. cerevisiae* invertase that was expressed and secreted by *P. pastoris* was demonstrated in two ways [14]. Retention of some of the endo H-released oligosaccharides on anion-exchange resin (quaternary aminoethyl-Sephadex) was attributed to negatively charged phosphate groups. Also, metabolic labelling of cell cultures with ^{32}P and subsequent isolation of the invertase-associated oligosaccharides revealed the presence of radioactivity in

the hyperglycosylated $Man_{30}GlcNAc$ fraction and in the $Man_{10-14}GlcNAc$ fraction. Alkaline phosphatase treatment of these oligosaccharides did not release phosphate. Thus it was concluded that the phosphate was not mono-esterified. Similar techniques and results were reported for the recombinant head domain of Influenza neuraminidase [34]. In addition, acid hydrolysis of the neuraminidase-linked oligosaccharides released small amounts of glucose phosphate and mannose phosphate, as demonstrated by high-performance anion-exchange chromatography [21]. For recombinant chicken α-*N*-acetylgalactosaminidase, it was suggested from MS data of tryptic glycopeptides that one phosphate group was present on some of the $Man_{9-14}GlcNAc_2$-bound structures [35]. But for the $\alpha_E C$ domain of human fibrinogen 420, it was suggested from mass-spectral analyis that the one N-glycosylation site contained approximately one-third of phosphorylated $Man_2GlcNAc_2$, the remainder being $Man_{6-13}GlcNAc_2$ (non-phosphorylated?) [24]. A comparative study of the N-linked oligosaccharide structures on six hydrolytic-type enzymes that were expressed in *P. pastoris* revealed some unique features [17]. In particular, on only one (aspartic protease) of the expressed proteins were phosphorylated oligosaccharides detected. Changes in HPLC and FACE mobilities after mild acid hydrolysis and alkaline phosphatase treatment suggested that the phosphate was present in diester form between mannose residues on $Man_{10-11}GlcNAc_2$.

We have analysed the acidic fraction of oligosaccharides that are assembled on the recombinant kringle 2 domain of human tissue-type plasminogen activator, when expressed in *P. pastoris* [29]. All N-linked oligosaccharides were released with PNGase F and the pool of fluorescently labelled oligosaccharides fractionated on an amino-silica column into a family of neutral and a family of charged components. Analysis of the charged components by DE-TOF-MALDI MS revealed masses that corresponded to individual major components from $Man_{10-14}GlcNAc_2$, plus 80 mass units, suggestive of a phosphate group being present. Treatment of these components with alkaline phosphatase did not alter the masses, indicating that the phosphomonoester linkage was not present. Treatment with mild acid hydrolysis followed by alkaline phosphatase reduced the masses by 132 mass units for a hexose and 80 mass units for a phosphate group, as expected for a phosphodiester linkage between an anomeric hydroxy and a primary (or secondary) hydroxy of two mannose units. The pool of observed charged oligosaccharides would then correspond to the $Man_{9-13}GlcNAc_2$ neutral oligosaccharides previously characterized, but each with an additional esterified mannose-phosphate group. The site of linkage of this group to the oligosaccharide has not been characterized, but may be identical to that found in other yeasts. The sites of phosphorylation that have been identified in *S. cerevisiae* mannoproteins are on both the $Man_8GlcNAc_2$ core and outer side chain of N-linked oligosaccharides, and on the short O-linked oligosaccharides [7]. Mannose 1-phosphate is transferred to carbon 6 of the central mannose residue in an acceptor sequence of Man(α-1,2)Man(α-1,2) or (α-1,6)Man, and such transfer appears to compete with addition of an α-1,3-Man residue to the non-reducing Manα-1,2 residue of the sequence. As α-1,3-mannosyltransferase activity is lacking in *P. pastoris*, it is possible that relatively large quantities of phosphorylated oligosaccharides, perhaps both N- and O-linked, would be found on native and recombinant glycoproteins.

Conclusions

Heterologous proteins that are expressed by *P. pastoris* cells can be glycosylated on the Asn component of the Asn-Xaa-Thr/Ser consensus sequence and on Ser or Thr hydroxy groups to provide N-linked and O-linked saccharides, respectively. Studies on the structures of N-linked oligosaccharides are consistent with earlier conclusions that a common yeast-type core structure of $Man_8GlcNAc_2$, probably originating from a dolichol-linked $Glc_3Man_9GlcNAc_2$, is elongated on the 1,3 arm with a chain of α-1,6-linked mannose units. These, in turn, can be branched with α-1,2-linked mannose units. The most common N-linked oligosaccharide that is assembled on proteins is $Man_{14}GlcNAc_2$, implying that short α-1,6 extensions occur, as compared with long (> 50) chains in proteins from *S. cerevisiae*. Consistently, α-1,3-linked mannose units are found to be absent in *P. pastoris*-derived proteins. Phosphorylated oligosaccharides are present, probably as Man-1-P-6-Man-R phosphodiesters, but exact sites of phosphate linkage on the oligosaccharides have not been identified. The O-linked saccharides are generally short (< 5 residues) and contain only α-1,2-linked mannose units. Attempts to engineer alterations in glycosylation processing are in early development, but may lead to glycoproteins similar to their mammalian counterparts in the future.

References

1 Gemmill, T. R. and Trimble, R. B. (1999) Biochim. Biophys. Acta 1426, 227–237
2 Burda, P. and Aebi, M. (1999) Biochim. Biophys. Acta 1426, 239–257
3 Knauer, R. and Lehle, L. (1999) Biochim. Biophys. Acta 1426, 259–273
4 Herscovics, A. (1999) Biochim. Biophys. Acta 1426, 275–285
5 Dean, N. (1999) Biochim. Biophys. Acta 1426, 309–322
6 Lussier, M., Sdicu, A. M. and Bussey, H. (1999) Biochim. Biophys. Acta 1426, 323–334

7 Jigami, Y. and Odani, T. (1999) Biochim. Biophys. Acta **1426**, 335–345
8 Gorin, P. A. J. and Perlin, A. S. (1956) Can. J. Chem. **34**, 1796–1803
9 Kobayashi, H., Shibata, N. and Suzuki, S. (1986) Arch. Biochem. Biophys. **245**, 494–503
10 Strahl-Bolsinger, S., Gentzsch, M. and Tanner, W. (1999) Biochim. Biophys. Acta **1426**, 297–307
11 Romanos, M. A., Scorer, C. A. and Clare, J. J. (1992) Yeast **8**, 423–488
12 Cregg, J. M., Vedvick, T. S. and Raschke, W. C. (1993) Biotechnology **11**, 905–910
13 Tschopp, J. F., Sverlow, G., Kosson, R., Craig, W. and Grinna, L. (1987) Biotechnology **5**, 1305–1308
14 Grinna, L. S. and Tschopp, J. F. (1989) Yeast **5**, 107–115
15 Trimble, R. B., Atkinson, P. H., Tschopp, J. F., Townsend, R. R. and Maley, F. (1991) J. Biol. Chem. **266**, 22807–22817
16 Verostek, M. F. and Trimble, R. B. (1995) Glycobiology **5**, 671–681
17 Montesino, R., Garcia, R., Quintero, O. and Cremata, J. A. (1998) Protein Expression Purif. **14**, 197–207
18 Garcia, R., Rodriguez, R., Montesino, R., Besada, V., Gonzalez, J. and Cremata, J. A. (1995) Anal. Biochem. **231**, 342–348
19 Calera, J. A., Paris, S., Monod, M., Hamilton, A. J., Debeaupuis, J. P., Diaquin, M., Lopez-Medrano, R., Leal, F. and Latge, J. P. (1997) Infect. Immun. **65**, 4718–4724
20 Fierobe, H. P., Mirgorodskaya, E., Frandsen, T. P., Roepstorff, P. and Svensson, B. (1997) Protein Expression Purif. **9**, 159–170
21 Martinet, W., Maras, M., Saelens, X., Jou, W. M. and Contreras, R. (1998) Biotechnol. Lett. **20**, 1171–1177
22 Wang, M. L., Lee, L. S., Nepomich, A., Yang, J. D., Conover, C., Whitlow, M. and Filpula, D. (1998) Protein Eng. **11**, 1277–1283
23 Reverter, D., Ventura, S., Villegas, V., Vendrell, J. and Aviles, F. X. (1998) J. Biol. Chem. **273**, 3535–3541
24 Applegate, D., Haraga, L., Hertzberg, K. M., Steben, L. S., Zhang, J. Z., Redman, C. M. and Grieninger, G. (1998) Blood **92**, 3669–3674
25 Spraggon, G., Applegate, D., Everse, S. J., Zhang, J. Z., Veerapandian, L., Redman, C., Doolittle, R. F. and Grieninger, G. (1998) Proc. Natl. Acad. Sci. U.S.A. **95**, 9099–9104
26 Rydberg, E. H., Sidhu, G., Vo, H. C., Hewitt, J., Cote, H. C. F., Wang, Y. L., Numao, S., MacGillivray, R. T. A., Overall, C. M., Brayer, G. D. and Withers, S. G. (1999) Protein Sci. **8**, 635–643
27 Nilsen, S. L., DeFord, M. E., Prorok, M., Chibber, B. A. K., Bretthauer, R. K. and Castellino, F. J. (1997) Biotechnol. Appl. Biochem. **25**, 73–74
28 Miele, R. G., Nilsen, S. L., Brito, T., Bretthauer, R. K. and Castellino, F. J. (1997) Biotechnol. Appl. Biochem. **25**, 151–157
29 Miele, R. G., Castellino, F. J. and Bretthauer, R. K. (1997) Biotechnol. Appl. Biochem. **26**, 79–83
30 Liao, Y. F., Lal, A. and Moremen, K. W. (1996) J. Biol. Chem. **271**, 28348–28358
31 Tsujikawa, M., Okabayashi, K., Morita, M. and Tanabe, T. (1996) Yeast **12**, 541–553
32 Van den Steen, P., Rudd, P. M., Proost, P., Martens, E., Paemen, L., Kuster, B., van Damme, J., Dwek, R. A. and Opdenakker, G. (1998) Biochim. Biophys. Acta **1425**, 587–598
33 Duman, J. G., Miele, R. G., Liang, H., Grella, D. K., Sim, K. L., Castellino, F. J. and Bretthauer, R. K. (1998) Biotechnol. Appl. Biochem. **28**, 39–45
34 Martinet, W., Saelens, X., Deroo, T., Neirynck, S., Contreras, R., Jou, W. M. and Fiers, W. (1997) Eur. J. Biochem. **247**, 332–338
35 Zhu, A., Wang, Z. K. and Beavis, R. (1998) Arch. Biochem. Biophys. **352**, 1–8
36 Scorer, C. A., Buckholz, R. G., Clare, J. J. and Romanos, M. A. (1993) Gene **136**, 111–119
37 Rodriguez, M., Rubiera, R., Penichet, M., Montesinos, R., Cremata, J., Falcon, V., Sanchez, G., Bringas, R., Cordoves, C., Valdes, M. et al. (1994) J. Biotechnol. **33**, 135–146
38 Paifer, E., Margolles, E., Cremata, J., Montesino, R., Herrera, L. and Delgado, J. M. (1994) Yeast **10**, 1414–1419
39 Vozza, L. A., Wittwer, L., Higgins, D. R., Purcell, T. J., Bergseid, M., Collins-Racie, L. A., LaVallie, E. R. and Hoeffler, J. P. (1996) Biotechnology **14**, 77–81
40 Sadhukhan, R. and Sen, I. (1996) J. Biol. Chem. **271**, 6429–6434
41 Romero, P. A., Lussier, M., Sdicu, A. M., Bussey, H. and Herscovics, A. (1997) Biochem. J. **321**, 289–295
42 Zhu, X., Wu, S. and Letchworth, III, G. J. (1997) Vaccine **15**, 679–688
43 Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T. and Kivirikko, K. I. (1997) EMBO J. **16**, 6702–6712
44 Masure, S., Paemen, L., Van Aelst, I., Fiten, P., Proost, P., Billiau, A., Van Damme, J. and Opdenakker, G. (1997) Eur. J. Biochem. **244**, 21–30
45 Sim, B. K. L., OReilly, M. S., Liang, H., Fortier, A. H., He, W. X., Madsen, J. W., Lapcevich, R. and Nacy, C. A. (1997) Cancer Res. **57**, 1329–1334
46 Abdulaev, N. G., Popp, M. P., Smith, W. C. and Ridge, K. D. (1997) Protein Expression Purif. **10**, 61–69
47 Berstmayer, B., Pessara, U. and Wels, W. (1997) FEBS Lett. **407**, 63–68
48 Chen, Y. J. and Gonatas, N. K. (1997) Biochem. Biophys. Res. Commun. **234**, 68–72
49 Kang, H. A., Sohn, J. H., Choi, E. S., Chung, B. H., Yu, M. H. and Rhee, S. K. (1998) Yeast **14**, 371–381
50 Rodenburg, K. W., Kjoller, L., Petersen, H. H. and Andreasen, P. A. (1998) Biochem. J. **329**, 55–63
51 Heimo, H., Palmu, K. and Suominen, I. (1998) Protein Expression Purif. **12**, 85–92
52 Weiss, H. M., Haase, W., Michel, H. and Reilander, H. (1998) Biochem. J. **330**, 1137–1147
53 Vailes, L. D., Kinter, M. T., Arruda, L. K. and Chapman, M. D. (1998) J. Allergy Clin. Immunol. **101**, 274–280
54 Austin, A. J., Jones, C. E. and Heeke, G. V. (1998) Protein Expression Purif. **13**, 136–142

55 McAlister, M. S., Davis, B., Pfuhl, M. and Driscoll, P. C. (1998) Protein Eng. 11, 847–853
56 McAlister, M. S., Brown, M. H., Willis, A. C., Rudd, P. M., Harvey, D. J., Aplin, R., Shotton, D. M., Dwek, R. A., Barclay, A. N. and Driscoll, P. C. (1998) Eur. J. Biochem. 257, 131–141
57 Vandersall-Nairn, A. S., Merkle, R. K., O'Brien, K., Oeltmann, T. N. and Moremen, K. W. (1998) Glycobiology 8, 1183–1194
58 Richard, F., Robert, P., Remy, J. J., Martinat, N., Bidart, J. M., Salesse, R. and Combarnous, Y. (1998) Biochem. Biophys. Res. Commun. 245, 847–852
59 You, Y. H., Hefta, L. J., Yazaki, P. J., Wu, A. M. and Shively, J. E. (1998) Anticancer Res. 18, 3193–3201
60 Greenwald, J., Le, V., Corrigan, A., Fischer, W., Komives, E., Vale, W. and Choe, S. (1998) Biochemistry 36, 16711–16718
61 Zhu, X., Wu, S. and Letchworth, G. J. (1999) Vaccine 17, 269–282

Received 12 July 1999; accepted 9 August 1999